

UC Irvine

UC Irvine Previously Published Works

Title

Impact of left censoring on longitudinal analyses of dementia using electronic health record data

Permalink

<https://escholarship.org/uc/item/9b61276r>

Journal

Alzheimer's & Dementia, 18(S11)

ISSN

1552-5260

Authors

Jiang, Luohua
Shi, Yuxi
Conniff, Kyle R
et al.

Publication Date

2022-12-01

DOI

10.1002/alz.066471

Peer reviewed

Impact of left censoring on longitudinal analyses of dementia using electronic health record data

Luohua Jiang¹ | Yuxi Shi² | Kyle R Conniff¹ | Maria M. Corrada¹ |
Spero Manson³ | Joan O'Connell³

¹University of California, Irvine, Irvine, CA, USA

²University of California Irvine, Irvine, CA, USA

³University of Colorado Anschutz Medical Campus, Aurora, CO, USA

Correspondence

Luohua Jiang, University of California, Irvine, Irvine, CA, USA.

Email: lhjiang@hs.uci.edu

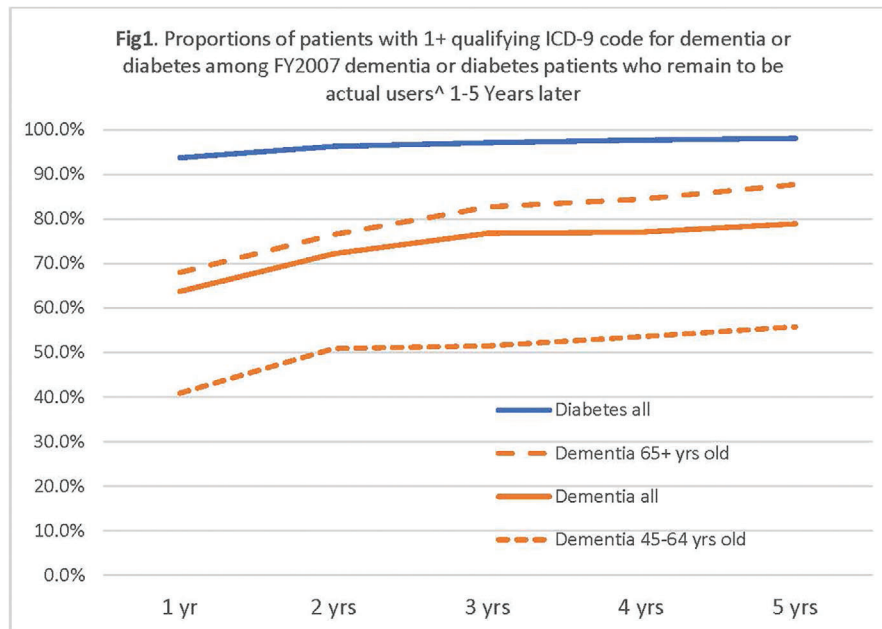
Abstract

Background: Longitudinal electronic health record (EHR) data provides an opportunity to evaluate associations between certain risk factors and dementia incidence. In practice, researchers using EHR usually extract data for multiple years and use the first 2-4 years as a washout period to exclude baseline prevalent cases. However, identifying dementia via clinical diagnostic codes during a washout period likely under-detect some cases, with time to dementia diagnosis for these cases left censored and estimates for the associations of interest potentially biased. This study aims to assess the amount of left censoring when analyzing time to dementia using EHR data from the Indian Health Service (IHS).

Methods: We extracted data from the IHS National Data Warehouse and related EHR databases between fiscal year (FY) 2007-2013. Adults aged 45+ years were identified as having dementia if they had at least one qualifying ICD-9 diagnostic code for all-cause dementia. A total of 1,117 AI/AN adults were identified as dementia patients in FY2007. Among these, we calculated the proportions of patients with 1+ ICD-9 code for dementia within 1-5 years after FY2007. We used diabetes, identified in a similar manner, as a comparison condition.

Results: Among the FY2007 dementia patients who used IHS services in FY2008, only 63.7% had 1+ qualifying ICD-9 code for dementia in FY2008. In contrast, 93.7% of FY2007 diabetes patients had an ICD-9 code for diabetes in FY2008 (Fig1). If FY2007 data are not included in the analytical dataset and only one year of data (FY2008) is used as the washout period, the remaining 36.3% prevalent dementia cases are classified as incidence after FY2008 with their time to dementia left censored. Even using a 5-year washout period, 25.0% prevalent dementia patients in FY2007 are classified as incident cases after FY2012 (Fig2a) while 4.6% prevalent diabetes patients in FY2007 are classified as incidence (Fig2b).

Conclusions: Longitudinal analysis of dementia risk using EHR data might be subject to relatively high amount of left censoring. It is important to recognize the potential biases associated with relatively high proportion of left censored observations when estimating longitudinal associations of risk factors with dementia incidence.



^ Indian Health Service (IHS) actual users in a fiscal year (FY) refer to those who used IHS services at least once with a diagnostic code in that FY.

